SOURCES OF ERROR IN CONTINGENT VALUATION*

Gary McClelland[1], William Schulze[2], Donald Waldman[2],
Julie Irwin[1], David Schenk[2]

December, 1990

[1] Department of Psychology and Center for Research on Judgment and Policy, University of Colorado-Boulder, Boulder, Colorado 80309

[2] Department of Economics and Center for Economic Analysis, University of Colorado-Boulder, Boulder, Colorado 80309

* This research was supported by the Office of Policy, Planning and Evaluation of the U.S. Environmental Protection Agency (USEPA) under Cooperative Agreement #CR-812054. We wish to hank Alan Carlin and Leland Deck of USEPA for comments. All opinions and conclusions are the sole responsibility of the authors.

DISCLAIMER

SOURCES OF ERROR IN CONTINGENT VALUATION

Gary McClelland, William Schulze, Donald Waldman,
Julie Irwin, David Schenk

1. Introduction

The use of surveys for valuing public goods became commonplace in the decade of the 1980's. The contingent valuation method (CVM) has, as a result of the need for damage assessments under the Comprehensive Environmental Response, Compensation and Liability Act of 1980 (CERCLA), now undergone Federal Court review. The decision of the court supports use of the method as a legitimate alternative to property value or other hedonic methods and to the travel cost approach in measuring natural resource damages. In great part, acceptance of the method has been based on a series of studies in which value estimates obtained by asking respondents for their willingness to pay (WTP) were compared to values obtained from indirect approaches such as the hedonic or travel cost method (see, for example Brookshire, Thayer, Schulze, d'Arge, 1982; or Smith, Desvousges and Fisher, 1986). In several field experiments actual purchase decisions have been compared to hypothetical purchase decisions (Bishop and Heberlein, 1978 and Dickie, Fisher, and Gerking, 1987). In all of these studies hypothetical behavior was sufficiently predictive of actual behavior that researchers concluded meaningful values could be obtained for benefit-cost analysis.

However, in their extensive review of the CVM literature, Cummings, Brookshire and Schulze (1986) note that in all of the available comparison studies, respondents necessarily had obtained at least some market experience with the commodity. For example, Brookshire, et. al. (1982) compared survey values obtained for air quality improvement in the Los Angeles area with air quality values obtained from a property value study. The premium found in the home sale market for areas with cleaner air is well known by area residents who experience a trade-off between housing costs and air quality in choosing where to live. Cummings, Brookshire and Schulze argue that studies of this type do not provide evidence that respondents have the ability to provide meaningful hypothetical values for public goods for which they have little or no prior market experience. By their very nature, many public goods do not allow market experience of the type obtained for air quality in the property value market described by Brookshire, et. al. For example, many studies using the CVM have shown large existence and bequest values for preserving environmental commodities (e.g., Greenley, Walsh and Young, 1981 and Schulze et. al. 1983). Freeman (1987) has argued that only the CVM can be used to measure those values since such preferences are not reflected in existing markets, denying both market experience and preventing use of indirect methods for valuation.

Hypothetical bias, defined as the difference between the distribution of hypothetical bids obtained from a survey and the distribution of bids that would obtain in a real world incentive compatible market setting, has thus become the central issue in application of the CVM. We argue that both lack of market experience and details of survey design (context) may contribute to hypothetical bias. In this paper we first identify four problem areas possible sources of hypothetical bias. These are: (1) large positive outlier bids; (2) refusals to bid; (3) viewing public goods are joint products and (4) survey context. The overall objective of this study is to examine these possible sources of hypothetical bias in a field application of the CVM.

The commodity chosen for the study, air quality in the Denver metropolitan area, has three features which make it appropriate for such an examination. First, a careful psychological study of how residents perceive air pollution in the region is available (Stewart et. al. 1983, 1984). Second, one of the primary features of Denver's air pollution problem, the "Brown Cloud" which obscures views of both the center city skyline and of the Colorado Front Range, is visible throughout the city. Thus, property value markets are little affected by air pollution, so residents have had little or no market experience with the commodity. Third, a high level of awareness of the problem and a community. consensus that something must be done has been achieved in the region. For example, the Chamber of Commerce has strongly supported new proposed air pollution controls and such innovative measures as currently required use of oxygenated fuels have received wide public support. Thus, although residents have had little or no market experience with the commodity, most have at least thought about the problem. Our choice of commodity can thus be seen as an attempt to examine hypothetical bias by moving away from market experience while still retaining a commodity for which the public has a clear sense of both the nature and importance of the commodity itself.

The remainder of the chapter is organized as follows: Section 2 discusses hypothetical bias. The design of the survey instruments used to value air quality is presented in Section 3. Section 4 presents results and data analysis and Section 5 provides conclusions and recommendations for future research.

2. Sources of Hypothetical Bias

Survey values obtained in the field tend to be bimodally distributed with a large number of zero bids and an upper mode which is skewed, showing a thick tail of large bids. Researchers have tended to view both the large number of zero and very high bids with considerable skepticism. Fortunately, laboratory experiments have shed considerable light on the problem of large bids which suggest a straightforward econometric solution. The problem of refusals to bid (either in the form of a stated but non-credible zero value or a non-response) creates problems both in identifying valid zero bids and of selection bias in estimating the true value of positive bids. Selection bias is a serious issue because a significant number of respondents in any CVM study may "conceal" their bids. Although procedures for dealing with selection bias are well known, we show that the interpretation of the results is problematic and they are sensitive to model specification. Another problem receiving increasing attention (see for example papers by Kahneman and Knetsch and by Smith, both forthcoming) concerns the motivation behind and content of bids obtained in CVM studies. We find that many respondents have a different view of the provision of public goods than that implicit in the way CVM questions are asked. They view an additional dollar of taxes as producing a variety of public goods as joint products. It is our view that this confusion (amongst CVM researchers, not respondents) has resulted in serious errors in interpreting responses. Finally if context effects are important, that is, if different survey designs obtain very different values for the same commodity, then estimated values are not robust. In this section we consider each of these problems in turn and suggest a specific set of procedures for each.

The Problem of Large Bids

Researchers have turned to laboratory economics experiments to understand the source of large hypothetical bids obtained in CVM studies. These Laboratory experiments typically place subjects in an unfamiliar environment (either with respect to the commodity, the market, or both) and compare an initial hypothetical response to actual laboratory market responses where repeated trials are used to provide market experience. We briefly summarize what has been learned from such experiments and, drawing on these experiments, propose both a specific model of hypothetical error and suggest an econometric approach for analysis of contingent values which may reduce hypothetical errors. Results from laboratory experiments show a consistent and striking pattern. Hypothetical bids obtained from subjects for a commodity show an increased variance relative to bids obtained in a laboratory market. Further, increasing market experience (repeated rounds in a particular auction institution) and increasing incentives (increased payoffs for participation in a particular market institution) both tend to reduce variance in bidding.

The first experiment to compare hypothetical bids to auction behavior, undertaken by Coursey, Hovis and Schulze (1987), used a bitter tasting liquid, sucrose octa acetate, which was unfamiliar to subjects as the commodity. Subjects were first given a careful description of the commodity and then were asked how much they would pay to avoid a taste experience. Second, subjects were allowed to taste the liquid prior to being asked again for their willingness to pay (WTP). In this second stage subjects were familiar with the commodity but had no market experience. Third, subjects participated in a competitive auction submitting bids to avoid the commodity. Mean bids (variance) were as follows: Hypothetical with no experience $2.60 ($15.80); hypothetical with experience with the commodity $2.27 ($5.06); and actual auction 'bids with market experience $1.95 ($5.23). Note, the variance is much greater for the inexperienced hypothetical bids. However it appears that the decrease in variance was associated with experience with the commodity rather than with experience with the market institution.

Other recent experiments which allowed more rounds of actual market experience than the Coursey, Hovis and Schulze experiment show a continued decline in bidding variance both with market experience and reward size (see Irwin, McClelland and Schulze, 1989 and Cox, Smith and Walker, 1989). Figure 1 shows how increasing variance in hypothetical bidding can bias estimates of actual behavior. The top panel of Figure 1 shows a skewed hypothetical distribution relative to the actual bidding distribution. The extended right hand tail is the source of an upward bias in the mean hypothetical bid. This source of error dominates the results of the experiments described above. The bottom panel shows a situation where an increase in hypothetical bidding variance would produce an unbiased estimate of mean bid.

Given the experimental evidence summarized above, what model can be used to explain hypothetical bias that might result in field surveys from a lack of, market experience? Assume for simplicity that all individuals have the same true willingness to pay, W. However, the bid they provide in response to a hypothetical question about willingness to pay is B. Then where ε is an additive hypothetical error (similar to that shown in the bottom panel of Figure 1) with a frequency distribution f(a),

$$B = W + \varepsilon \tag{1}$$

mean of B is then

$$\bar{B} = \int_{-\infty}^{\infty} f(\varepsilon)\,(W+\varepsilon)d\varepsilon$$

and is an unbiased estimate of W because we assume that hypothetical error is symmetrically distributed so that

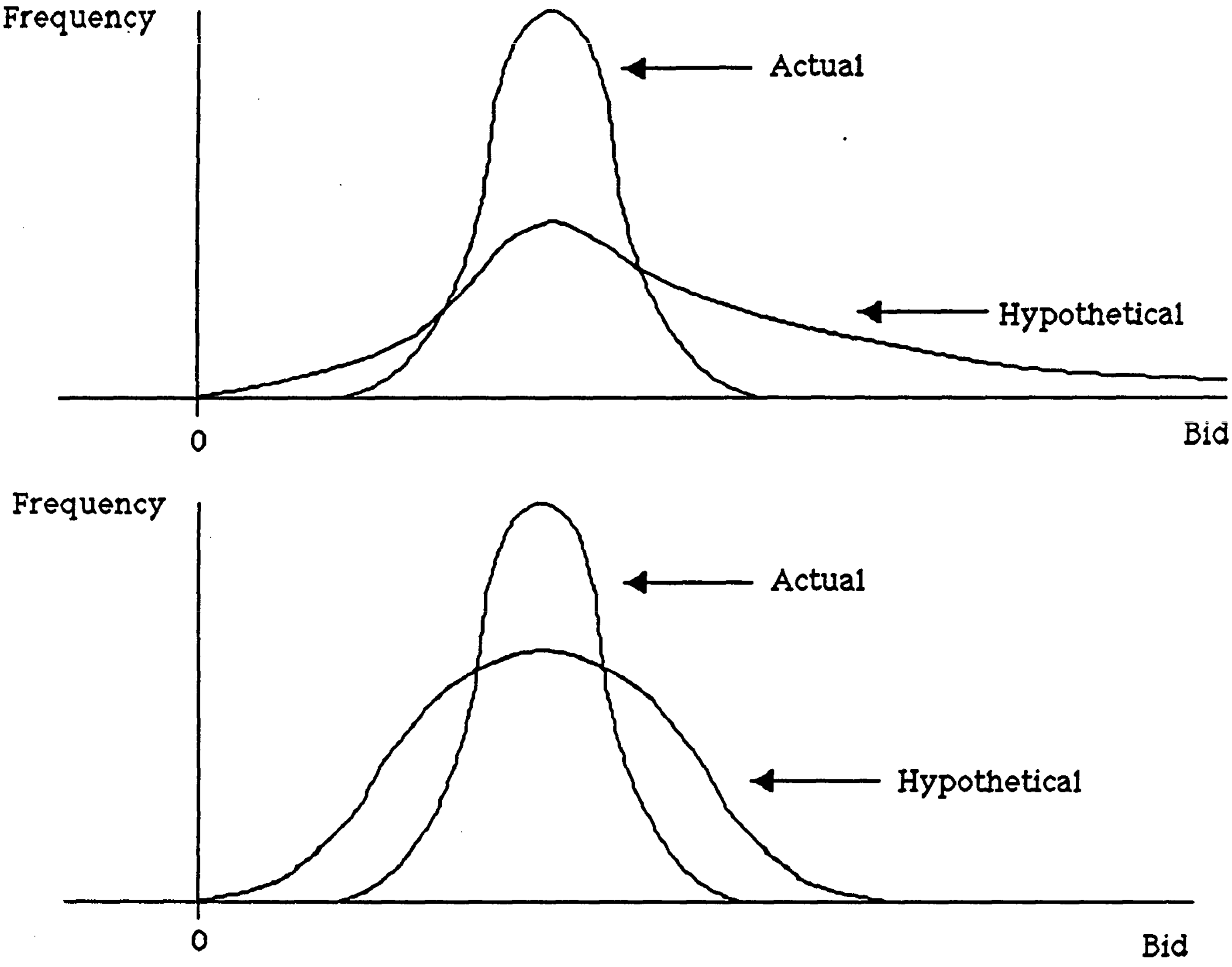
$$\int_{-\infty}^{+\infty} f(\varepsilon)\,\varepsilon d\varepsilon = 0. \tag{2}$$

In this situation use of the mean bid to estimate WTP will provide a true value unless a truncation bias occurs.

Obviously, this model cannot account for the problem of large bids. However, hypothetical error may not be symmetrically distributed. As noted above, available data both from the field and laboratory suggest that the error is highly skewed as is implied in the top panel of Figure 1. Assume for purposes of illustration that hypothetical error is distributed log normally so

$$\ln B = \ln W + \varepsilon.$$

Figure 1



The simple mean $\bar{B}$, will be upward biased, but the mean of lnB will be an unbiased estimator of lnW since (2) still holds. This implies that a transformation of survey bids (such as the natural logarithm) which produces a symmetrical error distribution may well reduce hypothetical bias by eliminating skew.

The intuition behind this model of proportional as opposed to additive hypothetical error is straightforward. Imagine that a respondent is first asked to hypothetically value an ice cream cone of a particular flavor which has a “true” value of $2.00 but makes errors on the order of ±$.50. Then the same respondent is asked to value a particular new car with a true value of $20,000. The question becomes, in valuing the new car is it plausible to assume that the error would also be on the order of ±$.50 (an additive error model), or on the order of ±$5,000 (a proportional error model)? In terms of the quick judgments asked for in contingent value studies, the proportional hypothetical error model is obviously more plausible. However, in a real situation a consumer would have much more incentive to find ways to reduce the size of error in judgment on the value of the car than on the value of the ice cream cone. This observation may, explain why initial large bids quickly fall in laboratory experiments. We propose use of a more general transformation, the Box-Cox, $(B^{\alpha}-1)/\alpha$, where α is determined to effectively normalize the error distribution in regression analysis (Box and Cox, 1964). It incorporates both the linear (α = 1) and natural logarithm (α = 0) transformations as possibilities. Use of this procedure has several advantages. In the past large suspect bids obtained in the CVM have been removed through trimming (e.g., Desvousges, Smith and Fisher, 1987). Trimming procedures remove large outliers which deviate from an estimated linear regression model by exceeding some predetermined statistical threshold. However, in the situation where the bid distribution shows a thick upper tail, the mean of predicted bids falls as that threshold is lowered, making final estimated values dependent on the threshold chosen. If skew is present, the procedure we propose will also lower mean values if bids generated by the estimated regression equation are used in calculating the mean. However, the reduction in predicted mean bid will be determined by the estimated value of α, the Box-Cox parameter, so as to make $f(\varepsilon)$ as normal as possible. If hypothetical error dominates the residual, then it is obviously desirable to use an estimating procedure which does not bias the estimated coefficients through a skewed hypothetical error distribution. Predicted values from this estimated equation can then be used to calculate mean or total willingness to pay.

Bid Refusals

A second problem in the interpretation and analysis of contingent values, which we feel is unresolved, is the presence of protest zero bids or refusals to bid when respondents are asked for willingness to pay (WTP). When pretesting survey instruments, researchers in debriefing respondents have often found that failures to bid or zero bids are not associated with a zero value to the respondent, but rather the respondent does not feel responsible for the problem and as a result conceals their bid. For example, a respondent may argue that although she is harmed by air pollution, she is not responsible for creating the problem (e.g., she does not own a car). Rather, industry and others should pay and are morally responsible for cleaning up the problem. Interestingly, such respondents when asked for their willingness to accept to allow a decrease in environmental quality often refuse any amount of money, arguing that to do so would be morally wrong. Thus, moral reasoning results in an unwillingness of respondents to provide any tradeoff between money and the public good in question. An apparent Lshaped indifference curve between money on the vertical axis and the public good on the horizontal axis results from the application of moral reasoning, a situation similar to that described by Hahneman (1989).

The mental process leading a respondent to conceal WTP (supported by debriefings in pretesting) may be as follows: “Cleaner air is very valuable to me so I would have to pay a lot to reflect that value; but, fortunately, air pollution is not my fault so I should not have to pay. I will bid zero or maybe I just will not answer the question since it does not apply to me.” Note that identification of such bid refusals can be accomplished in at least three ways in the design of a survey instrument all of which are employed in this study. First, a question asking why the respondent bid zero can be included. Second, a question asking for willingness to accept for a decrease in the level of provision can be included along with the WTP question. If the respondent indicates that no amount of money is morally acceptable as compensation for a reduction in provision, an associated zero WTP can be rejected as inconsistent. Third, questions asking how concerned about, bothered by, or important the commodity is to the respondent can be used to check for the consistency of a zero bid.

As argued by Smith and Desvousges (1987), the absence of positive bids from such respondents results in a potential selection bias problem since as many as one third of respondents may refuse to provide values. In estimating a regression model for those respondents who do provide a WTP value, selection bias must be accounted for to obtain unbiased coefficients (Heckman, 1979). However, as noted above, we find that although correcting for selection bias can have a large impact on predicted values, that the estimates are not robust to changes in model specification

Public Goods as Joint Products

Detailed debriefings of respondents obtained as part of pre-testing have often suggested that values are incorporated in WTP responses beyond those anticipated by investigators, who often struggle (unsuccessfully in our view) to limit bids to just the commodity or attribute under investigation. Our research suggests that some people have “mental models” of how the world functions which they maintain as a working hypotheses even in the face of “expert” evidence. These mental models usually relate to how governments raise and spend money. The investigator may promise that all the money raised by the program proposed for valuation will be spent on just the public good in question, but people with the view that governments produce services and commodities as joint products will bid for a vector of outputs rather than for the specific commodity. In pretesting the survey. instruments used here, we found, consistent with the arguments of Fischhoff and Furby (1987), that some respondents viewed health and visibility improvements as joint products. Further, some respondents viewed the proposed tax payments used to fund air quality improvements as providing for a much larger set of public goods. To formally explore these issues we use the following notation: Let

Q = Air Quality
v = Visibility
H = Healthiness of the Air
G = Other Public Goods
and X = Composite Commodity

If a respondent has a mental model such that

$$H = \alpha_H Q \quad V = \alpha_V Q \quad G = \alpha_G Q.$$

where α_H, α_V, and α_G are fixed coefficients, then the compensating variation measure of WTP for an improvement in visibility can be obtained by totally differentiating the constant level of utility of the consumer,

$$U(V,H,G,X^o\text{-}WTP) = U^o$$

subject to the joint product constraints listed above. The marginal willinness to pay for visibility then takes the form:

$$\frac{\partial WTP}{\partial V}\Big|_{U^o} = \underset{(a)}{\left(\frac{U_v}{U_x}\right)} + \underset{(b)}{\left(\frac{\alpha_H U_H}{\alpha_V U_X}\right)} + \underset{(c)}{\left(\frac{\alpha_G U_G}{\alpha_V U_X}\right)} .$$

Thus, if an individual who believes that government services are produced as joint products is asked to provide a bid for a small increase in visibility, the bid will contain marginal the willingness to pay for visibility (term a above), but will also contain appropriately proportioned values for related health improvements (term b above) and for increases in the provision of other public goods (term c above).

We test this hypothesis in two ways. First, in some survey variants we ask respondents to provide a dollar value for visibility improvement, then a separate dollar value for health improvements, and finally a total bid for the sum of visibility and health improvements for a specific air pollution program. Some people responded with three bids, bids in the following pattern: $50 for visibility, $50 for health, and $100 total. However, a large number of respondents gave bids in the following pattern, $100 for visibility, $100 for health, $100 total, consistent with the joint product hypothesis. A second way to examine this issue is to incorporate a debriefing question in the survey itself which asks respondents only for a total bid, but then asks them to split the bid up into its possible component parts. Thus, a respondent can plausibly state “my bid was 20% for visibility, 60% for health, and 20% for other public goods,” consistent with either preferences constrained or unconstrained by a joint product mental model.

Context

A potential problem in the design of any survey instrument is the degree to which the wording of the survey can affect respondents’ answers. Both the wording of the valuing question and the information surrounding the question, which we term the *context* of the question, can affect the value given. Hogarth (1982) in an edited volume presents a number of papers that confirm the notion that context can affect people’s responses, even in situations in which the context should logically have no effect. For example, researchers (e.g., Noell-Neumann, 1970) have found that the order in which (independent) questions are asked can affect people’s answers to the questions. Other researchers (e.g., Tversky and Kahneman, 1986, Lichtenstein and Slovic, 1971, 1973) have found that how the question is expressed (e.g., in terms of losses versus gains, or in terms of percentages instead of fractions) can affect people’s responses.

In order to understand context effects, it is helpful to think of values as being more or less *crystallized* (Schuman and Presser, 1981). If a person has had the opportunity to think about and/or obtain a choice experience with a commodity (in a marketplace or through a public decision process such as an actual referendum, for example) to such an extent that the value is “set” in the mind, then it is unlikely that the manner in which the value is elicited will affect the value. In such a case, we would say that the commodity’s value is *crystallized,* and relatively impervious to context effects. If, on the other hand, the commodity is one for which the person does not have a set value, because the commodity is not traded in a marketplace or has not been subjected to public debate and the person has not thought of the commodity in monetary terms, then the value for that commodity is less crystallized. In such a case, context difference could affect the value in two ways.

One way in which context can have an effect is in the process of evaluating the commodity. For example, when a respondent reads the words “air quality”, many components of the concept “air quality” may come to mind. In a sense, the words “air quality” themselves may have different meanings for different people, and for many people, the meanings will be quite vague and unformed. Context can help clarify the concept of air quality, or place emphasis on different aspects of the problem. Evidence has shown (Tversky and Kahneman, 1974) that whatever components of a commodity are most cognitively available to the respondent will figure most strongly in the evaluation of the commodity. For example, reading information about the health effects of particulates in the air in Denver could cause respondents to place more emphasis on the health dangers of air pollution, thus raising their values for improving air quality. If, on the other hand respondents hold strong opinions aout health effects, no impact of extra health information would be found.

Another type of context effect is rooted in the difficulty that people have in assigning values to commodities such as *air quality.* Context can help respondents understand how the general concept of "air quality" can be translated to a monetary scale, especially since it is likely that many respondents have not thought of environmental commodities in monetary terms. It is important, for that reason, that CVM questions be given enough context so that respondents believe and understand that their money would actually buy the commodity they are evaluating. Some researchers (Fischhoff and Furby, 1989) have suggested that even seemingly minor wording differences in CVM questions could result in respondents valuing essentially different commodities, thus making interpretation of the results impossible. The disadvantage to giving too much contextual information of this kind is the danger that the respondents will have a response to the contextual information that is unrelated to. their actual values for the commodity. For instance, asking respondents if they would be willing to be "taxed" to pay for "governmental programs" that would clean up the air pollution in Denver may result in angry refusals to give a value, not because the air pollution problem is unimportant to the respondents, but because of disgust for the government, or taxes, or some other element of the questions' context.

In order to test the degree to which values are crystallized, as well as the degree to which respondents need contextual information to make sense of CVM questions, we propose that context should be varied across different versions of any CVM survey. If values obtained across survey versions are robust with respect to changes in context, then it is likely that values are crystallized and hypothetical bias from this source may not be a severe problem for the particular commodity. In this study we test for context effects by implementing seven different designs of the survey which in some versions (1) vary the way the improvement in air quality is described, (2) provide additional information describing the health effects of air pollution, (3) use a referendum format as opposed to a straightforward WTP question, and (4) ask respondents to value other private commodities before valuing air quality.

3. Survey Design and Implementation

This section contains descriptions of the general design of the seven survey versions (see Table I for a summary of the survey designs), first outlining the common elements of the seven versions, and then listing and describing the elements that varied across survey versions.

The introductory sections of the surveys were similar across version: respondents indicated how bothered they were by the Denver Brown Cloud (the common local name for the air quality problem in Denver), where they had heard about the Brown Cloud problem, and so on. Following this set of questions, each survey asked respondents to look at a color photo insert. The insert contained 6 pictures of the Denver area under different visibility conditions and a seven-step ladder with two example photos anchoring steps 2 and 6. The photos used to anchor the ladder had been previously rated in pre-testing. Respondents used the 1-7 scale to rate each of the six

TABLE I

SUMMARY OF BROWN CLOUD SURVEY DESIGN FEATURES

		BASE		THREE		VOTING	FREQ. DIST.	3 COMMODITY COMPARISON
	VERSION	A	B	C	D	E	F	G
RESPONSE FRAME	WTP	X	X	X	X	X	X	X
	WTA	X	X			X		
HEALTH vs. VISIBILITY	3 Questions			X	X			
	% Split	X	X					
FORM OF THE VALUE QUESTIONS	Std. CVM	X	X	X	X		X	X
	Voting					X		
DESCRIPTION OF CHANGE IN AIR QULAITY	Average Air Quality Change	X	X	X	X	X		X
	Freq. Distribution of Air Qual. Change						X	
INFORMATION CONTENT	Health Information		X		X			

example photos, first in terms of visibility, and then in terms of healthiness of the air.

The next section asked respondents to think about the current overall air quality on an average winter day in the Denver metro area and to provide a rating for this average day (winter is the poorest season for air quality in Denver). The differing valuation questions followed and the last section of the survey, the demographic section, was the same for all versions. In this section, respondents were asked their age, income level, and so on.

The variables that define the different survey versions (as listed in Table 1) are explained and described below:

Response Frame

WTP All the survey versions contained a willingness to pay (WTP) question. The WTP question first asked respondents if they would consider paying for a one-step improvement in air quality, (whether this improvement was for visibility, health, or both depended on survey version). If they were not willing to pay, they were asked why; and if they were willing to pay, they were asked for a maximum dollar amount. For all of the WTP questions, this was a yearly household payment.

WTA Respondents to versions A, B, and E were also asked a willingness to accept compensation (WTA) question. The WTA question was identical to the WTP question, except that the WTA version asked respondents if they would be willing to accept compensation for a decrease in air quality, and the least they would be willing to accept monetarily to allow such a decrease.

Health vs. Visibility

Superadditivity In surveys C and D respondents gave WTP values for health, visibility, and total air quality. Respondents to the other survey versions gave values for air quality as a whole. This format explores the notion that the sum of health and visibility values will exceed the total value provided, consistent with the notion of a joint product mental model described above.

% Split In versions A and B, after responding to the WTP question, respondents also indicated the percent of their total air quality WTP value that they would attribute to health effects and the percent that they would attribute to visual air quality.

Form of Value Questions

Standard CVM Versions A, B, C, and D have WTP/WTA questions that follow the standard CVM procedure, with no added context.

Voting In accordance with the recommendations of Mitchell and Carson (1989) and of Fischhoff and Furby (1989), version E included added context for the WTP/WTA questions, containing more specific information about the connection between the WTP/WTA values and improvements/reductions in air quality. Instead of simply asking for payment/compensation for changes in air quality, the version E scenarios detailed two referenda (Mitchell and Carson, 1989) that would affect air quality as well as taxes and/or prices for which respondents could vote. Respondents in the WTP question are asked if they would consider voting for a referendum. If they answer “yes” they are asked what is the most they would pay before changing their mind.

Description of change in air quality

Average air quality change All of the versions except for version F asked respondents to value a one-step average increase/decrease in air quality on the air quality ladder.

Frequency distribution of air quality change Version F asked respondents to give WTP values for two types of increases in air quality, as expressed in frequency distributions of the number of days in a winter of bad medium and very good air quality levels. One frequency distribution showed an increase in very good days, with no change in bad days; the other frequency distribution showed a decrease in very bad days with no change in the good days. Each shift in the frequency distribution corresponded to a one-half step increase in the air quality ladder.

Context/Information Content;

Health information Versions B and D contained information on the probable chemical make-up (carbon monoxide, sulfur dioxide, and ozone) of the air depicted in the eight pictures (the six rated plus two example pictures) on the color insert. The information was scaled in accordance with the U.S. Environmental Protection Agency's Office of Air Quality Planning and Standards' (OAQPS) "Air Pollution Index", so that respondents would know the extent each day violated air quality standards on the three chemical variables. Also, versions B and D provided descriptions of the negative health effects of violations of these standards.

Three Commodity Comparison Survey version G, unlike the other survey versions, asked respondents to value two private commodities before valuing to air quality. This version asked respondents to consider a trade up of one step in air quality (which, except for the use of the word “trade”, is the same valuation in all the air quality WTP questions across versions), as well as a trade up from a less expensive to a more expensive camera and television set. The survey booklets contained photos depicting the less desirable and more desirable air qualities, TV’s, and cameras. The particular commodities and photos were chosen using pilot data, so that, on average across respondents, the camera trade would be slightly lower and the TV trade would be slightly higher in value than the air quality trade.

Survey Mailing

Surveys were designed consistent with the Dillman Total Design Method (Dillman, 1988). The Dillman recommendations involve personalized mailing including a hand-signed cover letter, hand-stamped envelopes, follow-up reminder postcards, and a second mailing to households that did not respond to the survey following the first mailing. Also, the surveys were printed and folded into a booklet measuring 8 inches by 6 inches. The surveys were six (versions A and C), seven (versions B and F), eight (versions D and E) or 9 (version G) pages long, including the cover and space for comments.

Each household in the sample received a version of the survey, a color photo insert, a cover letter, and an addressed, stamped envelope to return the survey to us. For the first mailing only, the package also included a two dollar bill, to thank respondents for their time and to encourage them to fill out and return the survey. In past survey research we have found that monetary incentives increase response rates significantly (Doane, et al., 1989).

The survey sample consisted of 1400 households (200 for each of the seven survey variations) randomly selected from the Denver Metro area, including the city of Denver and its neighboring suburbs of Arvada, Aurora, Littleton, and Wheat Ridge.

Response Rates

When corrected for bad addresses, the response rates ranged from 68.3% (version E) to 72.6% (version A). There was no statistically reliable difference in response rate by version (c^2 (6) = .42, ns.). The overall response rate was 71%, when corrected for bad addresses.

4. The Data, Econometric Model and Empirical Results

Variables

WTP. The WTP values used for comparison across versions are the single total value given for the WTP for a 1-step improvement in air quality in versions A, B, E, and G, the sum of the values given for visibility and health for a l-step improvement in the surveys with component value questions (versions C and D), or the sum of the two frequency distribution questions in version F. The frequency distribution questions each asked for a separate one-half step increase in air quality, so the sum of these two produced a WTP value comparable to the WTP values from the other versions ignoring any diminishing marginal utility of air quality improvement.. Although some respondents were asked for WTA as well as WTP values, only WTP is used for statistical analysis because so few respondents gave finite WTA responses. Out of the 342 respondents who returned versions of the survey with the WTA question, the vast majority indicated that they would not accept any amount of money for a decrease in air quality (i.e., their WTA value was “infinite"). Only 28 (8.2%)

respondents gave finite values for WTA. Out of the 812 survey responses, 283 (35%) of those respondents did not give positive WTP values. Three (1.06%) of these respondents indicated that they would not give a positive WTP value because air quality had no value for them (these responses were discarded for the Box-Cox analysis). The rest of the respondents either stated that they were not responsible for the problem or gave no reason. If more respondents had given zero dollar WTP bids for air quality, with the indication that the bid represented an actual zero value, then it would have been appropriate to include these zero dollar bids as valid responses. However, given that nearly all of the WTA values were "infinite" it is reasonable to conclude that, for many of the respondents, the refusal to bid in the WTP case indicated some sentiment other than an actual zero value for air quality. Thus, we define a dummy variable indicating if the respondent revealed a value (1) or concealed their value (0).

Three types of independent variables were used in the regressions: (a) survey design variables, (b) sociodemographic variables, and (c) air quality rating variables.

Survey Design Variables Dummy codes were used to indicate the design features of each survey version. The variable HEALTH denotes whether or not the survey version contained the additional health information supplied by OAQPS; Versions B and D contained this information. The value of 1 for the variable SUPERADD indicates that the survey version contained separate WTP questions for visibility and health effects. (Versions C and D). The variable VOTING denotes Version E which framed the WTP question in terms of a referendum. The variable FREQ indicates Version F which displayed the improvement in air quality in terms of changes in frequency distributions. Finally, the variable COMMOD indicates Version G which asked the WTP question for air quality after asking similar questions for two other commodities--a camera and a television.

Sociodemographics Standard sociodemographic variables included AGE, GENDER (1 = male, 2 = female), EMPLOYment (0 = unemployed, 1 = employed), EDUCation and INCOME. In past research on responses to risks, an important variable is whether or not there are children living in the household. The variable KIDS indicates whether or not there are any children 18 years or younger living in the household. MEDIA is an index describing how many times the respondent had read, seen or heard about the Brown Cloud in the media.

Air Quality Ratings Several questions asked respondents to evaluate the typical air quality they experienced and how much the visibility and health problems bothered them. VIS is the rating on a seven-point scale of "how bothered have you been by the Brown Cloud's effect on what you can see in the distance (mountains, buildings, etc.)?" HEAL is the rating on a seven-point scale of "have you or your family been bothered by any health problems which you believe to be caused or aggravated by Denver's air pollution?" In both cases higher ratings mean the respondent was more bothered. Finally, AVERT is the respondent's rating of the "current overall air quality (thinking of both health and visibility) on an average winter day" on the air quality ladder defined by the photographs enclosed with the survey.

<u>The Model</u>

The econometric model consists of two equations. The first equation determines the value of the environmental action. It is reasonable to use two, distinct equations in this setting because the situation is different from the conventional Tobit model. In that model the desire to make a positive bid would always result in a positive bid, so that a single equation accurately models both positive and zero bids. In the situation at hand, a person may indeed have positive value for cleaner air, but still bid nothing.

Therefore let y^*_{1i} be an index of the i-th persons desire to reveal his value of clean air. A linear model for y^*_{1i} is

$$y^*_{1i} = \underline{\beta}_1'\underline{x}_i + \varepsilon_{1i} \tag{4.1}$$

where $\underline{x}_i$ is a vector of explanatory variables $\underline{\beta}_1$ a vector of unknown parameters, and ε_{1i} a random disturbance. This index is not observed, of course, so define the indicator variable

$$y_{1i} = \begin{cases} 1 & \text{if } y^*_{1i} > 0 \\ 0 & \text{if } y^*_{1i} \leq 0 \end{cases} \tag{4.2}$$

The second equation determines true WTP, y^*_{2i}:

$$y^*_{2i} = \underline{\beta}_2'\underline{x}_i + \varepsilon_{2i} \tag{4.3}$$

where $\underline{\beta}_2$ is a vector of unknown parameters to be estimated and ε_{2i} is a random disturbance that is potentially correlated with ε_{1i}. Everyone places a value on an improvement in air quality (which may be positive, negative or zero), but only some choose to reveal it. Observed WTP is given by

$$y_{2i} = \begin{cases} y^*_{2i} & \text{if} \quad y^*_{1i} > 0 \\ 0 & \text{if} \quad y^*_{1i} \leq 0 \end{cases} \tag{4.4}$$

To complete the specification of the model, define $\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i})'$. This disturbance vector is assumed to be bivariate normally distributed with zero mean vector and variance-covariance matrix

$$V(\varepsilon_i) = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}, \tag{4.5}$$

and to be uncorrelated across observations. The variance of ε_{1i} has been normalized to one, since the scale of y^*_{1i} is not observed. The sign and size of σ_{12}, the covariance between e_{1i} and e_{2i}, are important in determining the effect of non-revealed bids on estimates of β_2, and in the estimation of WTP for those who have not revealed their true values.

Estimation Objectives

There are two objectives in estimation. The first is to determine the effects of changes in context and socio-demographic characteristics on WTP. Knowledge of these effects are important both practically, in aiding the design of surveys, and in equity considerations for policymakers designing abatement legislation and how to pay for it.

The second objective is to determine the true WTP for those who fail to reveal it. That is, we are interested in the WTP of all members of the sample, not just those who have decided to reveal it. Because of this second objective, we briefly sketch the well-known two-step estimation procedure (see Heckman, 1976) to set up the notation for what is to come.

Correcting for Samole Selection

To achieve these objectives the parameters of the second equation must be estimated. Straightforward linear regression of WTP on $\underline{x}_i$ would produce biased estimates of β_2, since, if the entire sample were used the observed zeroes are not true values, and if only positive WTP is used the sample is potentially nonrandomly selected. This can be seen by examining the conditional expectation of y^*_{2i}, conditional on positive values:

$$E(y^*_{2i}|y^*_{1i} > 0) = \underline{\beta}_2'\underline{x}_i + E(\varepsilon_{2i}|y^*_{1i} > 0) \; . \tag{4.6}$$

If the last term on the right hand side of the equation is not zero, estimates of $\underline{\beta}_2$ will be biased. Defining f and F as the density and distribution function of a standard normal random variable, this expectation can be written

$$E(\varepsilon_{2i}|y^*_{1i} > 0) = E(\varepsilon_{2i} | \varepsilon_{1i} > -\underline{\beta}_1'\underline{x}_i) = \sigma_{12}\overset{+}{\lambda}_i \; , \tag{4.7}$$

where

$$\overset{+}{\lambda}_i = \overset{+}{\lambda}_i(-\underline{\beta}_1'\underline{x}_i) = \frac{f(-\underline{\beta}_1'\underline{x}_i)}{[1 - F_{(-\underline{\beta}_1,\underline{x}_i)}]} \; . \tag{4.8}$$

This suggests adding an estimated of $\overset{+}{\lambda}_i$ to (4.3) to produce

$$y_{2i} = \underline{\beta}_1'\underline{x}_i + \sigma_{12}\hat{\overset{+}{\lambda}}_i + v_{2i} \; , \tag{4.9}$$

where v_{2i} has been implicitly defined to be equal to $\varepsilon_{2i} - \sigma_{1\,2}\hat{\overset{+}{\lambda}}_i$. This new disturbance will have zero expectation and be uncorrelated with $\underline{x}_i$ in large, samples.

The first step, then, is to apply probit to equations (4.1) and (4.2) to estimate $\overset{+}{\lambda}_i$. In the second step, equation 4.9 is estimated on the subsample of positive bids with $\hat{\overset{+}{\lambda}}_i$ replacing $\overset{+}{\lambda}_i$.

Probit Results

Results from this step are reported in Table 2. The context variable FREQ is marginally significant at conventional levels, but the other context variables are not important in determining whether WTP is revealed.

For the self-reported air quality rating variables, those who indicated larger amounts for the variable AVERT tended not to report their WTP, while the results were mixed and inconclusive on VIS and HEAL. For the sociodemographic variables, older individuals were generally less willing to reveal their bids, while those with higher incomes and more education were more willing.

The goodness of fit of the first step probit estimation has important implications for the second step (see Meng and Waldman, 1990). The likelihood ratio test statistic for the null hypothesis that the slope coefficients are zero is equal to 44.7, significant at a very low alpha level. Most of those who reported WTP were predicted to report (449/472), but only a few non-reporters were predicted not to report their WTP (32/223).

Willingness to pay regressions

The second step is the estimation of the WTP equation on the subsample of those willing to reveal their bids. The Box-Cox transformation was applied to WTP. That is, WTP was transformed according to the following formula:

$$WTP^* = \frac{WTP^{\alpha} - 1}{\alpha} .$$

In addition to the usual parameters of a regression model, a search was made for the value of a that maximized the fit. This is a well-known transformation that assumes the linear form when $\alpha = 1$ and approaches

Table 2
Probit Coefficients
Acceptance Equation (n=688)

Variable	Coefficient	T-ratio
cnst	.552210	.868
HEALTH	-.120085	-.880
SUPERADD	.107513	.783
VOTING	-.407537E-01	-.223
FREQ	.266481	1.517
COMMOD	.358296E-01	.209
VIS	.474569E-01	1.126
HEAL	-.365106E-01	-1.203
AVERT	-.180697	-3.296
GENDER	.956260E-01	.813
EMPLOY	.267317E-01	.591
AGE	-.938255E-02	-1.979
INCOME	.441257E-01	2.516
EDUC	.733450E-01	2.092
KIDS	.403669E-02	.036
MEDIA	-.500009E-01	-.836

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

		Predicted	
Actual	TOTAL	0	1
TOTAL	688	62	626
0	223	34	189
1	465	28	437

Chi-Squared (15)............ 46.992
Significance Level.......... .36927E-04

the natural logarithm as a approaches 0. We ignore the truncation problem raised by Poirer (1978) and others. These results are contained in Table 3. Three sets of results are reported. Coefficient estimates for the full specification are reported in the third and fourth columns. To analyze the effect of ignoring sample selection, the first column omit λ_i^+. Allowing for sample selection but not for survey design produces the results in the firth and sixth columns. Using the full model for WTP presented in columns' three and four of Table 3, we find some surprising results, First, HEALTH, VOTING, FREQ and COMMOD are all not significant at conventional levels. This suggests that context has little or no role in determining values in this particular situation i.e., values are robust to changes in context. The SUPERADD variable, denoting the summation of values, for health and visibility is however highly significant and positive, consistent with our joint product hypothesis. Also consistent with the joint product hypothesis is the fact that respondents who were asked to split their total value were readily able to do so (27% to visibility, 48% to health and 25% to other). Similarly, if we replace the observations on summed WTP with their own reported total WTP for the respondents in the SUPERADD treatment in the dependent variable (WTP across surveys) then the SUPERADD dummy variable becomes insignificant. Thus, we conclude that respondents can provide consistent and reasonable total values, and can split those values into % components, but cannot provide component dollar values because some respondents appear to view government programs as providing joint products. Note that such mental models do not reflect household production, but rather reflect how respondents view the external world.

Table 3

Box-Cox Regression Coefficients
With and Without Sample Selection Correction*
and Context Variables

Dependent Variable: WTP   n=465

	w/o lambda		All variables		w/o context	
Variable	**Beta**	**t**	**Beta**	**t**	**Beta**	**t**
cnst	3.08	2.28	5.40	3.41	6.91	4.66
HEALTH	-0.30	-0.91	-0.02	-0.06		
SUPERADD	1.30	3.95	1.12	3.34		
VOTING	-0.19	-0.42	-0.03	-0.08		
FREQ	0.85	2.14	0.52	1.27		
COMMOD	0.55	1.39	0.50	1.25		
GENDER	0.30	1.07	0.11	0.39	0.10	0.34
EMPLOY	0.01	0.06	-0.02	-0.23	0.06	-0.56
AGE	-0.01	-1.04	0.01	0.81	0.01	1.05
INCOME	0.23	5.45	0.15	2.97	0.12	2.57
EDUC	0.16	1.92	0.01	0.09	0.05	-0.56
KIDS	-0.14	-0.54	-0.11	-0.41	0.14	-0.53
MEDIA	0.05	0.41	0.15	1.13	0.17	1.25
lambda	----	----	-3.55	-2.76	4.48	-3.98
R-squared	0.13		0.15		0.12	

Box-Cox parameter = 0.163.

*See Table 2 for probit model determining sample selection correction.

The coefficient on $\hat{\lambda}_i^+$ is statistically significant, indicating that selection bias is important in this model. The negative sign on this variable means that unobserved factors influencing an individual to reveal his bid also cause that bid to be lower. The importance of controlling for selection bias can be seen by examining the first column, where λ_i^+ has been omitted. Note that the coefficients and t-statistics of FREQ and INCOM are considerably larger, and, even more strikingly, education appears to have a large and statistically significant effect on WTP. This compares to the first column where, when selection bias has been controlled for, the education variable is seen to be unimportant.

When the survey design variables are omitted, the income variable falls by approximately one-third and is not as precisely measured.

Predicting WTP

Table 4 contains descriptive statistics for willingness to pay. The first row is for the reported values, including zero when a bid was not revealed. The second row considers only the positive bids.

The next three sets of three rows each contain values of WTP predicted from the model for three subsamples: for the entire 688 individuals; for the 465 who revealed their bids; and for the 223 individuals who did not reveal their bids.

Rows 3-5 report predictions from the estimation without the sample selection correction (the coefficients reported in the first two columns of Table 3). Rows 6-8 predict WTP values from coefficient estimates from the full model (columns 3 and 4 of Table 3). In each case, these predictions are referred to as the unconditional predicted values. They answer the question “What would we predict for the value of clean air for an

Table 4
Statistics for WTP

	Mean	Std Dev	Minimum	Maximum	n
Observed bids:					
all*	155.33	325.90	0.00	5500.00	688
revealed	229.34	374.59	0.00	5500.00	465
Predicted values, Box-Cox transformation only:					
all	97.48	39.86	22.02	235.80	688
concealed	88.39	39.72	22.02	235.80	223
revealed	101.84	39.23	26.23	224.17	465
Unconditional predicted values, sample selection model with Box-Cox transformation:					
all	277.62	76.75	89.73	557.40	688
concealed	267.24	77.84	89.73	557.40	223
revealed	282.60	75.80	113.78	556.41	465
Conditional prediction of bids, sample selection model with Box-Cox transformation:					
all	277.62	76.75	89.73	557.40	688
concealed	1067.19	487.48	329.89	3130.36	223
revealed	121.86	52.29	14.32	310.63	465

* average includes 223 zero bids

individual, chosen at random from the population, with characteristics $\mathbf{x}_i$." The formula for this calculation is simply

$$E(y^*_{2i}) = \underline{\beta}_2'\underline{x}_i$$

The last two rows report predicted bids (as distinct from values), conditioned on the knowledge that it is known whether or not the individual would reveal their bid. They answer the question "What is the predicted bid (response to a question) for an individual, chosen at random from the population, with characteristics $\mathbf{x}_i$ for whom we know they would reveal their bid (4.10a) and for whom we know they would conceal their bid (4.10b). The formulas for these calculations are, respectively,

$$E(y^*_{2i}|y^*_{1i} > 0) = \underline{\beta}_2'\underline{x}_i + E(\varepsilon_{2i}|\varepsilon_{1i} > -\underline{\beta}_1'\underline{x}_i) = \underline{\beta}_2'\underline{x}_i + \sigma_{12}\overset{+}{\lambda}_i \tag{4.10a}$$

$$E(y^*_{2i}|y^*_{1i} \leq 0) = \underline{\beta}_2'\underline{x}_i + E(\varepsilon_{2i}|\varepsilon_{1i} \leq -\underline{\beta}_1'\underline{x}_i) = \underline{\beta}_2'\underline{x}_i + \sigma_{12}\bar{\lambda}_i \tag{4.10b}$$

where $\overset{+}{\lambda}_i$ is defined in equation 4.8 and

$$\bar{\lambda}_i = -\frac{f(-\underline{\beta}_1'\underline{x}_i)}{F(-\underline{\beta}_1'\underline{x}_i)} = -\frac{f(\underline{\beta}_1'\underline{x}_i)}{1 - F(\underline{\beta}_1'\underline{x}_i)} = -\hat{\lambda}_i^{+}(\underline{\beta}_1'\underline{x}_i)$$

where the second equal sign is based on the symmetry of the standard normal distribution and the third equal sign follows the definition of $\hat{\lambda}_i^{+}$ from equation 4.8 (see Maddala, 1983, p. 367). These formulas may be calculated with estimated values for $\underline{\beta}_1$ (and hence estimates of $\overset{+}{\lambda}_i$ and $\bar{\lambda}_i$)

from the probit estimation of equation 4.2, and estimates of β_2 and σ_{12} from the least squares (or Box-Cox) estimation of equation 4.9.

Since the estimate of σ_{12} is negative (-3.55), λ_i^+ strictly non-negative, and λ_i^- strictly non-positive, the three expectations in equation 4.10 will be ordered

$$E(y^*_{2i}|y^*_{1i} > 0) < E(y^*_{2i}) < E(y^*_{2i}|y^*_{1i} \leq 0) .$$

The reason for this ordering stems from the additional information about bids that is gained from the two equation model. The determination of the bid (equation 4.3) has two components: one deterministic, $\beta_2'\underline{x}$, and one random, ε_2. The random component includes all factors that might influence the bid that cannot be measured. In a single equation model, this disturbance would be assumed to be white noise, that is, nothing would be known about it. In the two equation model, some of those same unobserved factors might influence the desire to reveal one's bid. These are captured in the error in that equation, ε_1. Because there is information on both the size of revealed bids and who did not reveal their bids, the correlation between these two errors can be estimated. Thus, for each observation, the additional information on ε_2 is the knowledge that ε_1 was large (a bid revealer) or small (a bid concealers). The estimated measure of this information is precisely $\hat{\sigma}_{12}\hat{\lambda}_i^+$ for revealers and $\hat{\sigma}_{12}\hat{\lambda}_i^-$ for concealers.[2]

The predicted bid, then, is the sum of two components: the observed component, measured by $\hat{\beta}_2'\underline{x}_i$), and the unobserved component, $\hat{\sigma}_{12}\hat{\lambda}_i^+$ for

---

[2] This use of information about model errors in prediction is unusual, but not unique. Another situation where this occurs is in the one-step-ahead forecast of a time series with first-order autoregressive disturbances. Gee Goldberger, 1962.

revealers and $\hat{\sigma}_{12}\hat{\lambda}_i^-$ for concealers. The observed component for bid revealers is larger, on average, than for bid concealers. The reason for the large predicted bid of the concealers is due to the unobserved component. The decomposition is given by

	Extimated Bid		observed		unobserved
All observations:	$277.62	=	$277.62		
Revealed bids:	$121.86	=	$282.61	-	$160.75
Concealed bids:	$1,067.19	=	$267.24	+	$799.95

Another way to conceptualize what this means is to imagine outside-of-sample prediction. Suppose an individual is chosen at random from the population, and the information concerning income, education, etc. is recorded. The predicted value of that individual's WTP is $277.62, all from $\underline{\beta}_2' \underline{x}_i$. Now suppose that in addition to the demographic information, is is also (somehow) known that his individual would reveal his bid. Then his predicted bid would be reduced to $121.86. This is because the likely value of that individual's error is $ -160.75. Similarly, if it were known that he would conceal his bid his predicted bid would be increased by the expected error in that case, $799.95.

5. Conclusions

We postulated four sources of possible hypothetical error in Section 2 and have examined, the implications of each in a field study using seven alternative survey designs. While we can draw preliminary conclusions from one application, we have of course, raised more questions than we have answered.

<u>Large Bids</u>

The Box-Cox procedure we proposed for positive bids (as a replacement for trimming) has considerable appeal in that it has support both from laboratory experiments and econometrics where a normal distribution of residuals is highly desirable. The Box-Cox coefficient obtained in our analysis, about.16, is close to the logarithmic transformation supported by psychological arguments of proportional error. The implications of this procedure can be obtained by examining the raw mean and the predicted mean for the Box-Cox model (ignoring sample selection) for revealed bids in Table 4. The raw mean bid is $229.34 in contrast to the predicted mean bid of $101.84. Thus, the Box-Cox procedure has a dramatic impact on values.

<u>Bid Refusals</u>

The issue of the treatment of refusal bids remains problematic. Application of the selection bias model on the assumption that the zero or non-bidders are concealing their values raises estimated values above the raw means as is readily apparent in Table 4, Until a better understanding of bid refusal is obtained, we would advocate caution in applying the selection bias model. This caution is further supported by the observation that if we include the self-reported (and possibly endogenous) air quality variables in the WTP regression of the full model, λ becomes insignificant and predicted values are similar to those obtained by using the Box-Cox procedure alone.

Public Goods as Joint Products

The initial tests of the joint product mental model we conducted all support both the hypothesis itself and the need for the use of debriefing questions to determine what people themselves have included in their bid. Such questions are being used with considerable success (see for example Chestnut and Rowe, 1990).

Context

Given the recent emphasis on the importance of context effects in application of the CVM, we find the non-significance of context in this application to be quite surprising. However, we caution that other commodities may not be robust to variations in context.

REFERENCES

Box, G.E.P.. and D.R Cox, (part 2), 1964. "An Analysis of Transformations," Journal of the Royal Statistical Society, Series B, XXVI, 211-243.

Bishop, RC., and T.A. Heberlein, 1978. "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?" American Journal of Agricultural Economics Vol. 61, No. 5, 926-930.

Brookshire. D.. M. Thayer W.D. Schulze. R d'Arge. 1982. "Valuing Public Goods: A Comparison of the Survey and Hedonic Approaches," American Economic Review, Vol. 72, No. 1., 165-177.

Cox, J.C., V.L. Smith, and J.M. Walker, 1989. "Reward Saliency, Dominance, and Decision Costs in Experimental Economics." Paper presented at Economic Science Association, Fall Meetings.

Coursey, D., J. Hovis, and W.D. Schulze, 1987. "On the Supposed Disparity Between Willingness to Accept and Willingness to Pay Measures of Value," Quarterly Journal of Economics, 679-690.

Cummings, R.G., D.S. Brookshire and W.D. Schulze, 1986. Valuing Environmental Goods: An Assesssment of the Contingent Valuation Method, Rowman & Allanheld, publishers.

Dickie, M., A. Fisher, and S. Gerking, 1987. "Market Transactions and Hypothetical Demand. Data: A Comparative Study," Journal of the American Statistical Association, Vol. 82, No. 397, 69-75.

Dillman, D.A., 1978. Mail and Telephone Surveys: The Total Design Method, John Wiley & Sons, N.Y.

Desvousges, W.H., V.K. Smith, and A. Fisher, 1987. "Option Price Estimates for Water Quality Improvements: A Contingent Valuation Study for the Monongahela River," Journal of Environmental Economics and Management Vol. 1 14:3,148-267.

Doane, M., G.H. McClelland, W.D. Schulze, C.K. Woo, 1989. 'The Value of Electrical Power Outages: A Contingent Valuation Study," Report submitted to Niagara Mohawk Electric Power Company.

Fischhoff B. and L. Furby, 1988. "Measuring Values: A Conceptual Framework for Interpreting Transactions with Special Reference to Contingent Valuation of Visibility, Journal of Risk and Uncertainty, forthcoming.

Freeman, M.A. III, 1987. "Non-Use Values in Natural Resource Damage Assessment," Manuscript, Bowdoin College.

Goldberger, A. (1962). "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," *Journal of the American Statistical Association*, 57:424-31.

Greenley, D.A., R.G. Walsh, and R.A. Young, 1981. "Option Value: Empirical Evidence from a Case Study of Recreation and Water Quality," Quarterly Journal of Economics Vol. 96. No. 4, 657-672.

Hanemann, M.W., 1989. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" Manuscript, University of California, Berkeley, California.

Heckman, J., 1979. "Sample Selection Bias as a Specification Error," Econometrica.

Heckman J. (1976). "The common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement* 5:475-92.

Hogarth, R.M, ed., 1982. Question Framing and Response Consistency. Jossey-Bass. Inc., Publishers. San Francisco, California.

Irwin, J.R, G. H, McClelland and W.D. Schulze, 1989. "Hypothetical and Real Consequences to Low Probability Risk," Submitted to Journal of Behavioral Decisionmaking.

Kahneman, Daniel and Jack L. Knetsch, 'Valuing Public Goods: The Purchase of Moral Satisfaction," forthcomig, JEEM.

Lichtenstein, S., and P. Slovic, 197 1. "Reversals in Preference Between Bids and Choices in Gambling Decisions," Journal of Experimental Psychology. 98, 46-55.

Lichtenstein, S., and P. Slovic, 1973. "Response-induced Reversals in Preference in Gambling: An Extended Replication in Las Vegas," Journal of Experimental Psychology, 101. 16-20.

Maddala, G.S., 1983. Limited-Dependent and Qualitative Variables in Econometrics Econometric Society Monographs, Cambridge University Press, Cambridge, Mass.

Meng, C. and D. Waldman (1990). "Alternative Estimators in the Evaluation of Voluntary Programs." Manuscript, University of Colorado.

Mitchell, RB. and RT. Carson, 1988. Using Surveys to Value Public Goods The Contingent Valuation Method Published by Resources for the Future, Washington, D.C.

Noelle-Neuman, E., 1970. 'Wanted: Rules for Wording Structural Questionnaries," Public Opinion Quarterly, 34, 191-20 1.

Poirer, D. (1978) 'The Use of the Box-Cox Transformation in Limited Dependent Variable Models," *Journal of the American Statistical Association*, 73, 284-7.

Schulze, W.D., et al., 1983. "The Economic Benefits of Preserving Visibility in the National Parklands of the Southwest," Natural Resources Journal, Vol. 23, 151-173.

Schulze, W.D., 1988. "Use of Direct Methods for Valuing Natural Resource Damages," Paper prepared for the Conference on Natural Resource Damage Assessment, June 15-17, 1988.

Schuman, H. and S. Presser, 1981. Questions and Answers in Attitude Surveys: Experiments on Question Form, Working and Context. Academic Press, Inc. Harcourt Brace Jovanovich, Publishers.

Smith, V.K., W.H. Desvousges, and A. Fisher, 1986. "A Comparison of Direct and Indirect Methods of Estimating Environmental Benefits," American Journal of Agricultural Economics , Vol. 68:2

Smith, V.K., and W.H. Desvousges, 1987. "An Empirical Analysis of the Economic value of Risk Changes," Journal of Policital Economy, Vol. 95, No. 1, 89-114.

Smith, V. Kerry, "Arbitrary Values, Good Causes, and Premature Verdicts: A Reation to Kahneman and Knetsch," forthcoming, JEEM.

Stewart, T. R, P. Middleton, and D. Ely, 1983. "Judgments of Visual Air Quality: Reliability and Validity, Journal of Environmental Psychology, 3, 129-145.

Stewart, T.R., P. Middleton, P., M. Downton, and D. Ely, 1984. "Judgments of Photographs vs. Field Observations in Studies of

Perception and Judgment of the Visual Environment. Journal of Environmental Psychology, 4, 283-302.

Tversky, A. and D. Kahneman, 1974. "Judgment Under Uncertainty: Heuristics and Biases," Science, 185, 1124-1131.

Tversky, A., and D. Kahneman, 1981. 'The Framing of Decisions and the Psychology of Choice," Science, 211, 453-458.

THIS PAGE INTENTIONALLY LEFT BLANK